4.1.4 Review and Correction via Data Analysis

Review and Correction via Data Analysis option of the REVIEW AND CORRECTION module allows you to conduct a search, then analyze the resulting data using graphs. This option is used to specify a certain range, then identify outlier cases that fall beyond that range and need further review. A link to the ID by Item screen allows you to mark cases to include or exclude from analysis or to mark items to estimate separately.

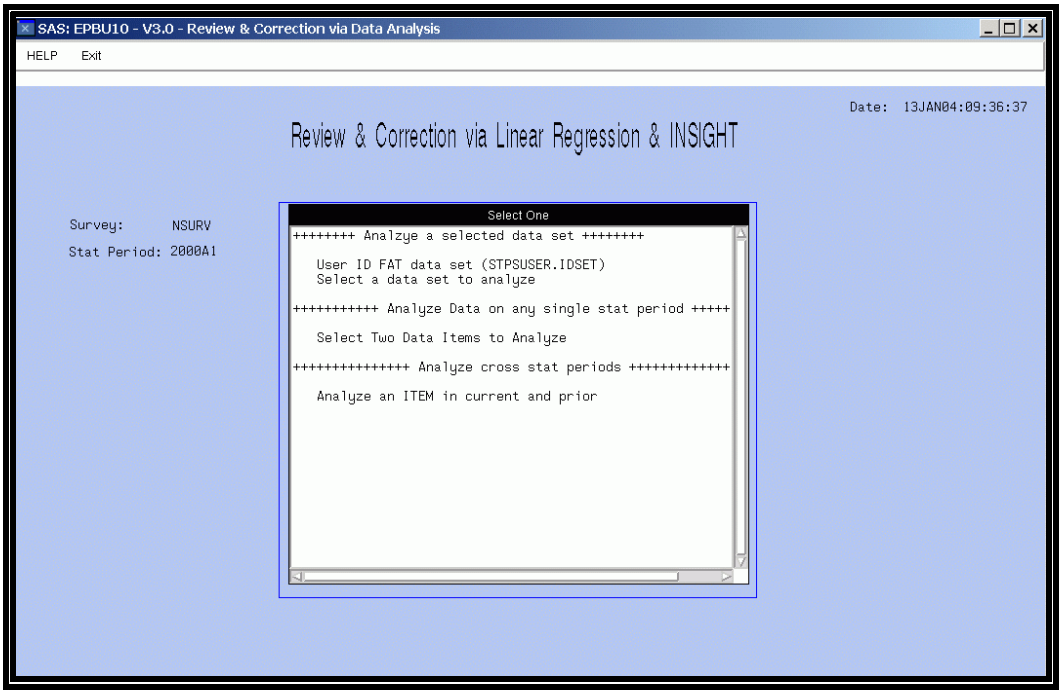

Figure 4.1.4a Initial screen for Review and Correction via Linear Regression & INSIGHT

The screen for this option, shown as Figure 4.1.4a, has the title "Review & Correction via Linear Regression & INSIGHT." To access this screen:

- To display this screen, click on the REVIEW AND CORRECTION BUTTON from the StEPS MAIN MENU
- Select Option 4 - Review and Correction via data analysis.

To use this screen, first click to select the type of data to analyze: a selected data set, data on any single stat period, or an ITEM that appears in both the current and prior stat period.

All options first display a "library" screen that lists the libraries and data sets or the fat record sets that are available for the survey and stat period selected via the User Setup button (see chapter 2).

When a PROCEED button is clicked the Data Browse screen is displayed. This section first describes "library" and Data Browse screens, which are common to all of the analysis options. It then describes each of the INSIGHT® options.

4.1.4.1 OVERVIEW OF SCREENS USED BY THIS OPTION

All three of the options on screen 4.1.4a use a variation of the screen that has a standard format for

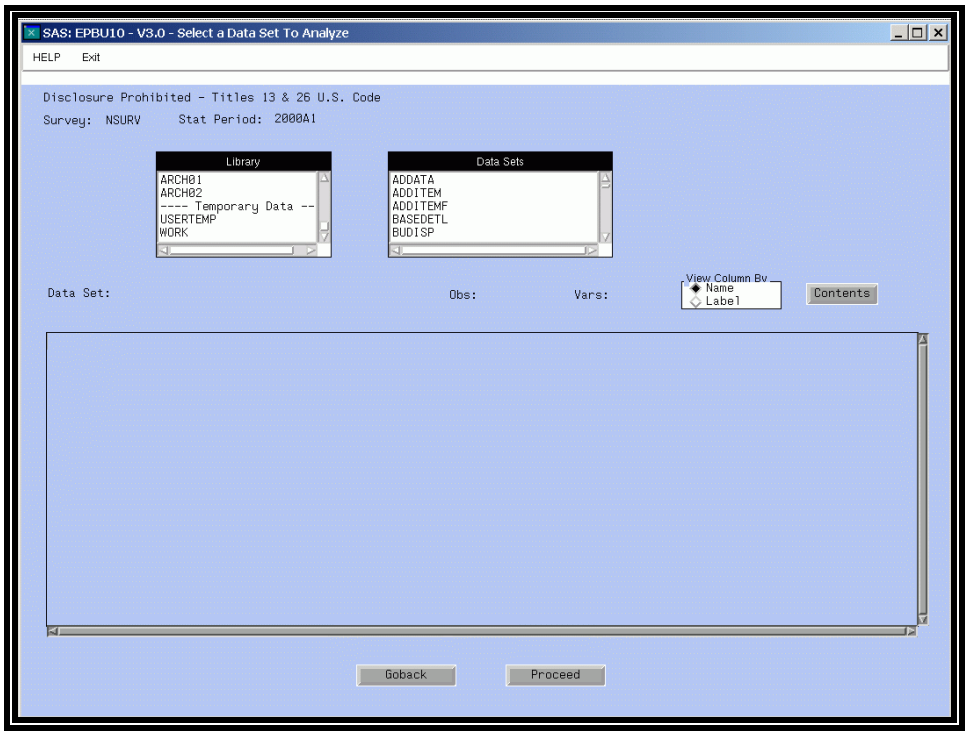

Figure 4.1.4b Select Data to Browse – screen showing available libraries and data sets

selecting a data set to browse (see Figure 4.1.4b for an example). When the data set is selected, it displays in the area at the bottom of the screen to allow you to verify your choice.

When click the PROCEED button, the data set will display in the Data Browse screen (see Figure 4.1.4c).

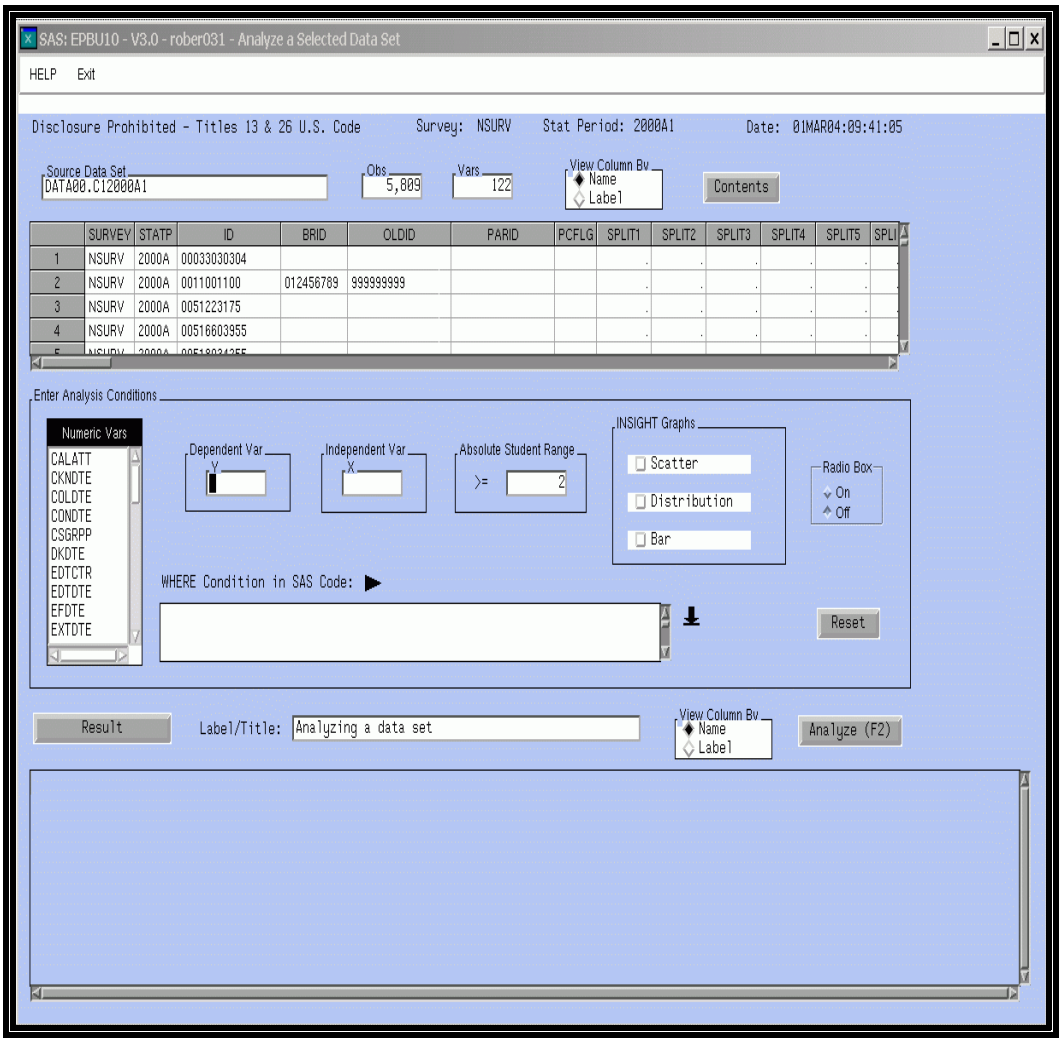

Figure 4.1.4c Data Browse screen

The top of the screen displays the data set (or fat record set) chosen.

The box labeled “Enter Analysis Conditions” varies by type of analysis chosen on the initial screen. These fields are described in detail under the appropriate analysis.

All of the options have a feature called “INSIGHT® Graph” displayed in a box on the right-hand side of the screen. This featurecreates a graph of the selected data. You may choose one or more of three graphing options: scatter, distribution, and/or bar graph.

- A field labeled “Radio box” controls the display of the graph. If Radio Box is “on,” the graph(s) will be displayed in a separate window. If this option is “off,” the graph(s) will not be displayed.

- When you click on ANALYZE or press F2, the selected data is analyzed and the results are displayed in a table. If a graphing option has been selected, a separate window will display the graph. If you select two or more graphic options, each will display in a separate window.

Figure 4.1.4d shows a screen after a data set and a graphic option have been chosen. The graph type chosen was HISTOGRAM. There are three windows displayed. One has the initial Data Browse screen. One has the data that was selected for the graph. One has the histogram. The details of selecting "x" and "y" variables, using range tests, and other analysis tools available through this option are explained in sections 4.1.4.2 through 4.1.4.4. Take the SAS INSIGHT® class offered at the Bureau for more information on using this graphical data analysis package.

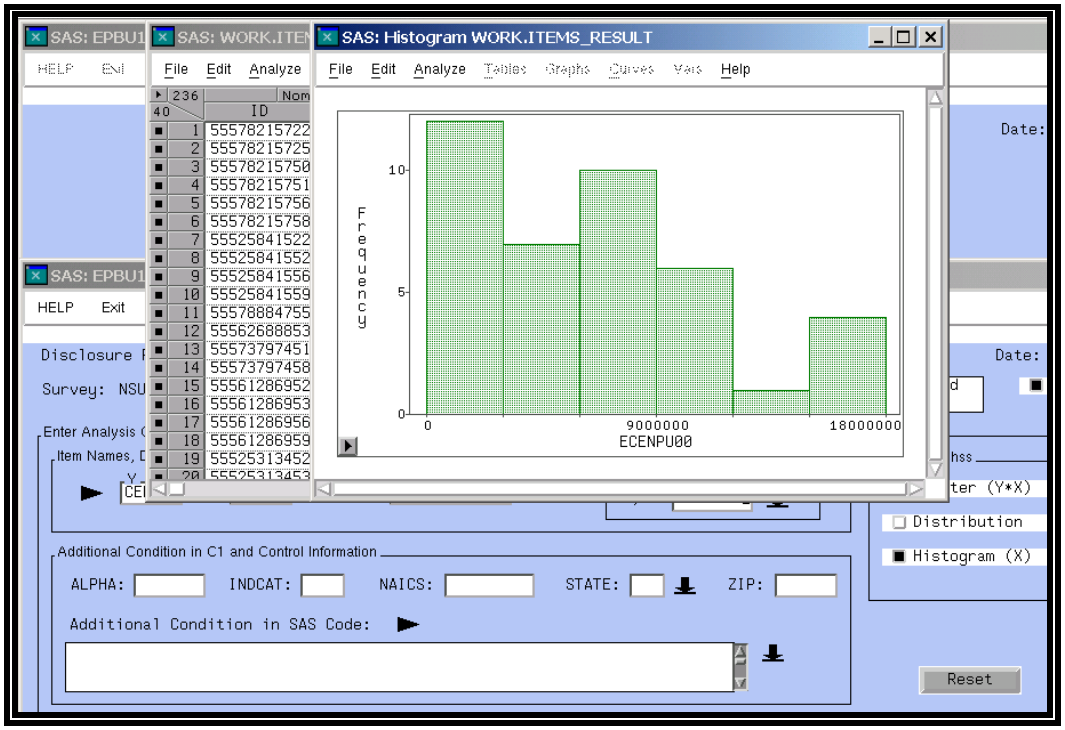

Figure 4.1.4d Example of histogram produced from selected data

If your analysis reveals problems with the data, you can use the Data Browse screen to mark IDs for inclusion or exclusion in the analysis, then rerun the graphing option.

- Go to the data set listing at the bottom of the screen (Figure 4.1.4b)
- Click on the ID you wish to examine

- The following options appear:

- Access “ID by Item” screen with this ID <xxxx>
- Access “ Control Data” screen with this ID <xxxx>
- Access “Respondent Text” screen with this ID <xxxx>
- ID – No Label
- Set Bookmark row <x>

When Access ID by Item is selected, The “ID by Item” screen appears with the selected ID at the top of the screen. To mark this item for action, click on the ID. The following options will appear:

- B - Bad data, exclude
- D - Different basis, excude
- I - Include, machine
- O - Item failed edit, exclude
- S - Estimate separately, exclude
- U - Failed edit, exclude
- Blank - Initial value, exclude

Accessing the ID by Item screen from the Data Browse screen is quite slow. If you want to make many changes, it is better to save the analysis as a user data set and go into the ID by Item screen from the StEPS Review and Correction via selection set option (see sections 4.1 and 4.2).

The Access Control Data option displays the Control Data screen for the ID. You can make corrections and save them (see section 4.6).

The Access Respondent Text screen displays any respondent text. Again corrections can be made (see section 4.12).

The “ID - No Label” field toggles. If a label exists, this line displays the label.

You can bookmark a row. If you click this line to bookmark the row, it will be added to a list of bookmarks that displays whenever you click this option. You can then click on the bookmark to return to this ID.

4.1.4.2 ANALYZE A SELECTED DATA SET

Option 1 - User ID Fat Data Set (STPSUSER.IDSET)

This screen allows you to analyze data from an ID Fat Data Set created using the TOOLS module (see detailed description in Chapter 5). Screen options are:

Source Data Set: This field shows you the source for the data set that is being viewed and analyzed.

Obs: This field displays the number of observations (rows) in the specific data set.

Vars: This field displays the number of variables (columns) in the specific data set.

"View Column By": Click on the box to choose to view columns either by the variable "Name" as it exists in the data set or by "Label."

Contents: Clicking on the "content" button will bring up a separate informational box that gives the general properties of the data set. This is information generated when the data set was created or last modified.

Numeric Vars: The box displays the numeric variables that are available within the data set. You can scroll down to view all possible variables and then select one. You are then prompted to either send that particular variable to "X" as the independent variable or "Y" as the dependent variable. (The "X" variable will appear on the horizontal axis of a graph and the "Y" on the vertical axis of the graph.)

Dependent Variable: This will be the "X" variable. You can either enter the dependent variable into this field or select one from the "numeric vars" box.

Independent Variable: The "Y" variable." You can either enter the independent variable into this field or choose a variable from the "numeric vars" box and designate that variable as the independent variable.

Absolute Student Range: This field is essentially a range test based on a calculated standard mean and standard deviation. Zero is the mean. For this test, you have to compute the value that represents a deviation that is outside the tolerance acceptable. You may enter any number. Any value that is greater than or equal to (>=) to the specified amount will be considered out of range. Both IDs that exceed the value on the positive side and those that fall below the value on the negative side will be shown as outliers.

For example, if you wanted to examine all IDs that had a large difference in value between edited data and adjusted data:

> Select EDDATA as the dependent variable (X axis)
> Select ADDATA as the independent variable (Y axis)
> In the Absolute Student Range, type in a number for the maximum percentage of a variance to be examined. If you type in 1000 and click ANALYZE (or press F2), a table will display all data items and associated IDs where values vary by more than plus or minus 1000 between the edited and the adjusted versions of the data. If there are no outliers, a message saying there is no outliers will be displayed.

Insight Graphs: As explained, this field allows you to select the type of graph (scatter, distribution, or bar) to display and analyze. Clicking in the box will activate that particular graph and clicking again will deactivate it. You may click on multiple graphs.

Radio Box: As explained in 4.1.4.1, if the radio box is "on" then the graph(s) will be displayed and if it is "off" the graphs will not be displayed.

WHERE condition in SAS code?: You may enter a WHERE clause here to specify parameters for the analysis. Clicking on the arrow above the field will give you a choice to either use a saved WHERE clause or to bring up the standard WHERE clause screen, which would assist you in creating a condition. By clicking on the arrow to the right of the box, you can choose to run the search, check syntax, clear WHERE, or save. For example: If you wanted to analyze only IDs that had been checked in, you would select CHKDTE from the list of variables, then select LE (equal to or less than) from the list of operators, then select "CONSTANT (enter value)" from the list of variables. When you select CONSTANT (enter value) a box will appear for you to type in the cutoff date you want. See section 4.1.1 a detailed explanation on constructing WHERE clauses.

Reset: Clicking on reset will clear the DEPENDENT VARIABLE, INDEPENDENT VARIABLE, and ABSOLUTE STUDENT RANGE fields and allow you to reenter information.

> **"View Column By"**: This field alters the column labels that are displayed and allows you to either view the information by "Name" as it exists in the data set or by "Label", which replaces the variable name with the descriptive label put in the SAS data set by the programmer. If no descriptive label exists, the column will just display the "Name."
>
> **Analyze (F2)**: Clicking on the "Analyze" button (or pressing F2) will process the selected information and analyze the desired data. The numerical data will be displayed in the bottom table while the graphs (if selected and the RADIO BOX set to on) will be displayed in separate screens.
>
> **Result**: This box will display the number of observations once a search has been performed. Clicking on this box with bring up the following options:
>
> 1. Save to a Permanent Data Set
> 2. Print Result – portrait
> 3. Print Result – landscape
> 4. Access Search/Extract/Utility Screen
> 5. Create ID Selection Set
> 6. Create Fat Record Data Set

Note: If no search has been performed and no cases are displayed, the "Result" button is inactive and the message "There is nothing to work on" will appear at the top of the screen in red.

Option 2 - Select a Data Set to Analyze

This option displays the StEPS Data Browse screen with the graphics options added. After selecting the library and data set, you can click on "Proceed" and will then be directed to the analysis screen described in Option 1. The only difference is that the information from the selected data set will be displayed and analyzed rather than the information from the Fat Record.

4.1.4.3 ANALYZE DATA ON A SINGLE STAT PERIOD

This option takes you to another version of the analyze screen and allows two data items to be entered and analyzed within a single stat period. The top portion of the screen displays the survey and stat period with the option of choosing two additional stat periods as well as options for searching the current or prior stat period and saving the search parameters upon exiting. These features are explained previously in 4.1.3 "Review and Correction via Searches." The rest of the screen features and analysis conditions include:

Item Names: There are two fields labeled "X" and "Y" where you can enter in the desired items for the search. There are arrows to the side of the fields and clicking on these arrows will produce a pick list of items from which you can choose. The "X" variable will be the horizontal axis and the "Y" variable will be the vertical axis of a graph.

Data Version: Choose the "Data Version" by clicking on the arrow and choosing either 1)Edited, 2)Reported, 3)Adjusted, or 4)Weighted. This option simply narrows the search according to which type of data you are looking for.

Absolute Student Range: This field allows you to enter a value that will represent distance from zero, essentially creating a range test. This instance of Absolute Student Range allows you to enter a value and provides a pick list to select a percentage between zero and 100. The analysis will listall cases that are equal to or greater than the selected percent. For example, to find all IDs where payroll is greater than a 40% of SALES:

Click on CTPAY (total payroll) and assigning it to the dependent axis
Click on STSAL (total sales) and assigning it to the independent axis
Click on the pick list by Absolute Student Range and select "40.".
Press F2 or clicks ANALYZE to display the results.

Insight Graphs: This field allows you to check boxes for scatter, distribution, and bar graphs to be displayed. Clicking in the box will activate that particular graph and clicking again will deactivate it. See the discussion of graphs at the beginning for section 4.1.4.

Additional Conditions using C1 and Control Information: This sections includes fields for entering in control information such as ALPHA, INDCAT, NAICS, STATE, and ZIP to narrow down the search. It also provides space to enter or create a WHERE clause. See section 4.1.2 for further details on entering control information and see section 4.1.1 for information on constructing a WHERE clause.

Reset: Clicking on reset will clear the x and y fields.

Label/Title: This field changes the column headings. If there is no descriptive label entered for the column, the title only will display.

"View Column By": This field alters the column labels that are displayed and allows you to either view the information by "Name" as it exists in the data set or by "Label", which replaces the variable name with the descriptive label entered by the programmer.

Analyze (F2): Clicking on this button (or pressing F2) will process the selected information and analyze the desired data. The numeric data is displayed at the bottom of the screen. Any selected graphs are in a separate window.

Result: This box will display the number of observations once a search has been performed. Clicking on this box with bring up the following options:

1. Save to a Permanent Data Set
2. Print Result – portrait
3. Print Result – landscape
4. Access Search/Extract/Utility Screen
5. Create ID Selection Set
6. Create Fat Record Data Set

Note: If no search has been performed and no cases are displayed, the "Result" button is inactive and the message "There is nothing to work on" will appear at the top of the screen in red.

4.1.4.4 ANALYZE CROSS STAT PERIODS

Selecting this option allows you to compare values for an item for the current stat period and the prior stat period. The screen is basically the same for this option as it is for "select two data items to analyze" except that there is only one item field and there is not an option for you to specify which stat period to search.

In the ITEM field where you can either enter an item or click on the arrow and select an item from the pick list. Above this field, it is stated that the independent variable, X, is the data in the prior stat period and the dependent variable, Y, is the data in the current stat period. You may then choose which data version to analyze by clicking on the arrow and choosing: 1) Edited, 2) Reported, 3) Adjusted, or 4) Weighted.

The field labeled "Absolute Student Range" has an the arrow to the right that provides a pick list of possible percent values, ranging from 0 to 100. This range test is described with an example in section 4.1.4.3.

The fields for Insight Graphs, Additional Condition in C1 and Control Information, Reset, Analyze (F2), Label/Title, "View Column By", and Result follow are described in the previous option, "Analyze Data on a single stat period".

P-Menus

P-Menu for Options 4 - Review and Correction via data analysis

P-Menu	Options	Function
HELP	Data Review and Correction Menu Help (F1) Function Key Help WhoamI (F7)	Display HELP information on using the Review and Correction Main Menu screen Display list of function keys and descriptions Display user default and systems information
EXIT	StEPS Main Menu (Home) Exit (F3)	Return to StEPS Main Menu Exit to previous screen